AHRQ

AHRQ QUALITY INDICATORS

# Prevention Quality Indicators

SOFTWARE DOCUMENTATION

**AHRQ Quality Indicators**

# Prevention Quality Indicators: Software Documentation, Version 2.1 – SPSS

Department of Health and Human Services
Agency for Healthcare Research and Quality
http://www.qualityindicators.ahrq.gov

October 2001
AHRQ Pub. No. 02-R0207
Revision 4 (November 24, 2004)

## Citation

AHRQ Quality Indicators -- Prevention Quality Indicators: Software Documentation, Version 2.1 – SPSS. Rockville, MD: Agency for Healthcare Research and Quality, 2001. Revision 4 (November 24, 2004). AHRQ Pub. No. 02-R0207.

## Abstract

The value of information on health care quality has never been so widely recognized; yet many organizations lack the resources and/or expertise to build a quality information program from the ground up. Recognizing this, the AHRQ Quality Indicators were developed specifically to meet the short-term needs for information on health care quality using standardized, user-friendly methods and existing sources of data.

This module of the AHRQ Quality Indicators was designed to capitalize on the availability of administrative data on inpatient stays to produce information about 16 Prevention Quality Indicators (PQIs): *potentially avoidable hospitalizations for ambulatory care sensitive condition* (ACSC) indicators, which involve admissions that evidence suggest could have been avoided, at least in part, through better access to high-quality outpatient care.

This document is the software documentation for the Prevention Quality Indicators software Version 2.1 (Revision 4), which is provided on the AHRQ Web site. The software was developed in SPSS, for use on a personal computer. By making this tool available, we hope to assist others in producing information on health care quality more cost effectively.

Details on the development of the Prevention Quality Indicators can be found in "Guide to Prevention Quality Indicators: Hospital Admission for Ambulatory Care Sensitive Conditions."

# Acknowledgments

This product is based on the work of many individuals who contributed to its development and testing.

The following staff from the **Evidence-based Practice Center (EPC) at UCSF-Stanford** performed the evidence review, completed the empirical evaluation, and created the programming code and technical documentation for the new Quality Indicators:

***Core Project Team***
Mark McClellan, M.D., Ph.D., *principal investigator*
Kathryn M. McDonald, M.M., *EPC coordinator*
Sheryl M. Davies, M.A.
Jeffrey Geppert, J.D.
Patrick Romano, M.D., M.P.H.
Kaveh G. Shojania, M.D.

***Other Contributors***
Amber Barnato, M.D.
Paul Collins, B.A.
Bradford Duncan M.D.
Michael Gould, M.D., M.S.
Paul Heidenreich, M.D.
Corinna Haberland, M.D.
Paul Matz, M.D.
Courtney Maclean, B.A.
Susana Martins, M.D.
Kristine McCoy, M.P.H.
Suzanne Olson, M.A.
L. LaShawndra Pace, B.A.
Mark Schleinitz, M.D.
Herb Szeto, M.D.
Carol Vorhaus, M.B.A.
Peter Weiss, M.D.
Meghan Wheat, B.A.

***Consultants***
Douglas Staiger, Ph.D.

The following staff from **Social & Scientific Systems**, **Inc.**, developed this software product, documentation, and guide:

***Programmers***
Leif Karell
Kathy McMillan
Fred Rohde

***Technical Writer***
Patricia Caldwell
***Graphics Designer***
Laura Spofford

Contributors from the **Agency for Healthcare Research and Quality**:
Anne Elixhauser, Ph.D.
H. Joanna Jiang, Ph.D.
Denise Remus, Ph.D., R.N.
Margaret Coopey, R.N., M.G.A, M.P.S.

Support efforts, including refinement and enhancement of the AHRQ Quality Indicators and related products, are provided by the support for Quality Indicators-II contract team.

The following individuals from **Battelle Memorial Institute, Stanford University,** and **University of California (UC)** constitute the Support for Quality Indicators-II core team:

Sheryl M. Davies, M.A.
Bruce Ellis, M.S.
Jeffrey Geppert, J.D.
Mark Gritz, Ph.D.
Greg Hubert, B.S.
Elaine Keller, M.Ed.
Kathryn M. McDonald, M.M.
Patrick Romano, M.D., M.P.H
Jeff Schoenborn, BS

The **Agency for Healthcare Research and Quality** Support for Quality Indicators-II team includes:

Denise Remus, Ph.D., R.N., Project Officer
Mary B. Haines, Contract Officer

We also wish to acknowledge the contribution of the peer reviewers of the evidence report and the beta-testers of the software products, whose input was invaluable.

## Table of Contents

**Abstract** ..... i

**Acknowledgments** ..... ii

**Introduction** ..... 1

**Components of the Prevention Quality Indicators Module** ..... 2

**Quick Reference** ..... 2
Prevention Quality Indicator Module Details ..... 5

**Background** ..... 7

**Data Elements and Coding Conventions** ..... 8
Input for the AHRQ Quality Indicators ..... 8
Coding of Diagnoses, Procedures, and DRGs ..... 9
Missing Values ..... 13
Treatment of Missing Data ..... 13

**User Options** ..... 14

**Processing Steps** ..... 15
1. Identify Outcomes in Inpatient Records ..... 15
2. Identify Populations at Risk from Census Population Data ..... 15
3. Calculate Observed (Raw) Prevention Quality Indicator Rates ..... 15
4. Risk Adjust the Prevention Quality Indicator Rates ..... 15
5. Create MSX Smoothed Rates ..... 16
6. Calculate Condition-specific Rates ..... 16

**Program Descriptions** ..... 17
Program 1: PQSPSA1.SPS ..... 17
Program 2: PQSPSA2.SPS ..... 18
Program 3: PQSPSA3.SPS ..... 20
Program 4: PQSPSC2.SPS ..... 21

**Reviewing the Printed Output** ..... 23
PQSPSA1.SPS ..... 23
PQSPSA2.SPS ..... 23
PQSPSA3.SPS ..... 24
PQSPSC2.SPS ..... 25

**Interpreting the Results** ..... 26
Observed Means ..... 26
Comparing Observed and Condition-Specific Means ..... 26
Risk-adjusted Rates ..... 26
MSX Smoothed Rates (Optional) ..... 27
Comparing Observed, Risk-adjusted and Smoothed Rates ..... 28

**Benchmark Timings** ..... 30

**User Support and Information** ..... 30

**Appendix A: Log of Revisions to PQI Documentation and Software Version 2.1, Revision 4** ..... A-1

**Appendix B: ICD-9-CM and DRG Coding Updates in PQI Release Version 2.1, Revision 4..... B-1**

## List of Tables

Table 1. Prevention Quality Indicator (PQI) Variables ..... 5
Table 2. Prefixes for the Prevention Quality Indicator (PQI) Variables ..... 5
Table 3. Prevention Quality Indicator Module Contents ..... 6
Table 4. Data Elements and Coding Conventions ..... 11
Table 5. Treatment of Missing Data Elements ..... 13
Table 6. PQSPSA2 Stratification Choices ..... 19
Table 7. Comparing Observed Rates with Risk-adjusted Rates ..... 29
Table 8. Comparing Risk-adjusted Rates with Smoothed Rates (Optional) ..... 29

## List of Figures

Figure 1 Processing Steps for the PQIs - Area ..... 3
Figure 2 Processing Steps for the PQIs - Condition-specific ..... 4

# AHRQ Quality Indicators, Software Documentation, Version 2.1: Prevention Quality Indicators

## Introduction

This documentation describes the software for implementing the Prevention Quality Indicator (PQI) module of the AHRQ Quality Indicators Version 2.1 (Revision 4) and highlights features of the analytic approach of particular interest to new users. Detailed information about the Prevention Quality Indicators (PQI) is contained in the companion document "*Guide to the Prevention Quality Indicators*[1]*,"* which includes information about the development of the indicators and their definitions.

The software consists of a series of computer programs that:

- Assign and calculate PQIs from hospital discharge abstracts or UB (uniform billing) claims data.
- Print PQI results in SPSS format that can be imported into word processing documents, spreadsheets, or graphics software, at the user's option.
- Provides the option to generate rates by Metropolitan Statistical Areas (MSAs) or by county for urban counties (rural areas are defined by county in either case).
- Calculate rates using either the hospital location or the county of patient residence.
- Create risk-adjusted rates that adjust for casemix differences (defined by age and sex).
- Create smoothed rates that reduce fluctuations over time due to random variation.
- Provides the option to generate condition-specific rates (diabetes) by state and age.

This documentation includes:

- A discussion of the data elements necessary to produce the AHRQ PQIs and the uniform coding conventions recommended for those elements.
- Descriptions of the PQI processing steps in nontechnical language.
- Detailed descriptions of the functions of each PQI SPSS program.

***Critical user information is highlighted with this symbol:*** 

[1] Available at http://www.qualityindicators.ahrq.gov/pqi_download.htm.

## Components of the Prevention Quality Indicators Module

As shown in Table 3 on page 6 the Prevention Quality Indicators module consists of three SPSS programs and 15 ASCII text files that contain auxiliary data. These programs and text files are described in the subsequent sections of this document. The programs were developed for use in Statistical Package for the Social Sciences[2] (SPSS), version 7.5 or higher, on an Windows personal computer.

A note on the naming conventions for the Prevention Quality Indicators (PQI) module programs. The programs have names of the form PQSPSAi.SPS. The first two characters "PQ" of the program name indicates a Prevention Quality Indicators program, to distinguish it from other modules that are available or forthcoming from AHRQ. The next three characters of the program name are "SPS" and are present to distinguish the program from the SAS versions of the software. The sixth character of the program name is either "A" to denote a program that is for the production of Area indicator rates that use county and/or Metropolitan Statistical Area (MSA) area populations as denominators, or "C" to denote a program for the production of condition-specific indicator rates.[3] The last character (i) designates the number of the specific program.

## Quick Reference

The subsequent three pages provide information intended to serve as a quick reference to assist in reading this documentation and in reviewing the Prevention Quality Indicator (PQI) module outputs. Processing steps (Figure 1 and Figure 2) are shown first followed by a listing of the module variables (Table 1), followed by variable prefixes (Table 2), and finally a listing of the SPSS module contents including the SPSS syntax files and text reference files used by the program (Table 3).

[2] SPSS is a statistical program distributed by SPSS, Inc. (http://www.spss.com). The company may be contacted directly regarding the licensing of their product. SPSS, Inc. does not have any affiliation with AHRQ nor involvement in the development of the AHRQ QIs.

[3] PQI version 2.1, revision 4 offers a new feature – the ability to produce condition-specific indicator rates for the four diabetes PQIs (PQIs 1, 3, 14, and 16). The user can calculate these PQIs based on the area population or by the prevalence of condition, i.e., the number of diabetics in a State stratified by age.

**Figure 1 Processing Steps for the PQIs - Area**

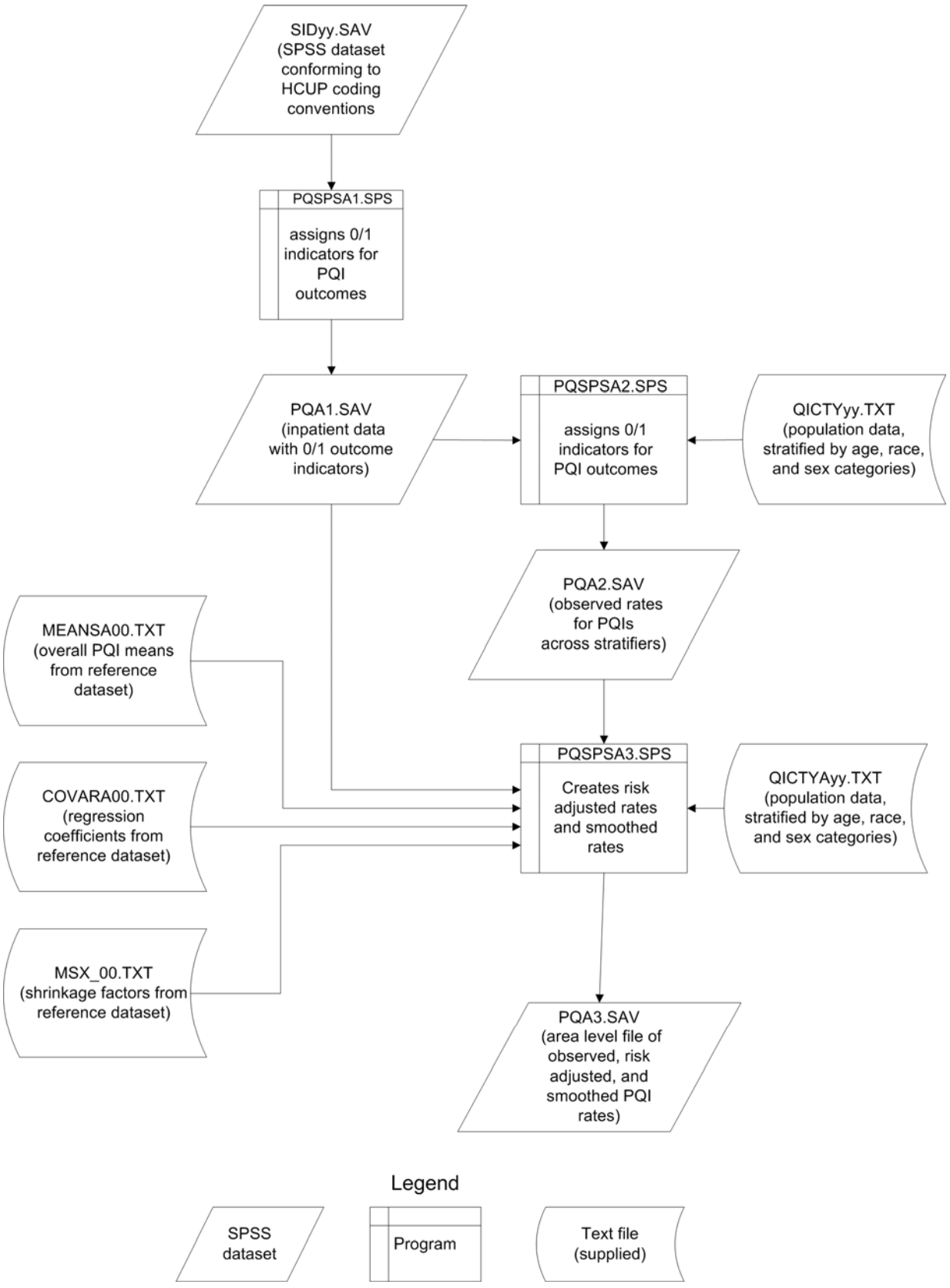
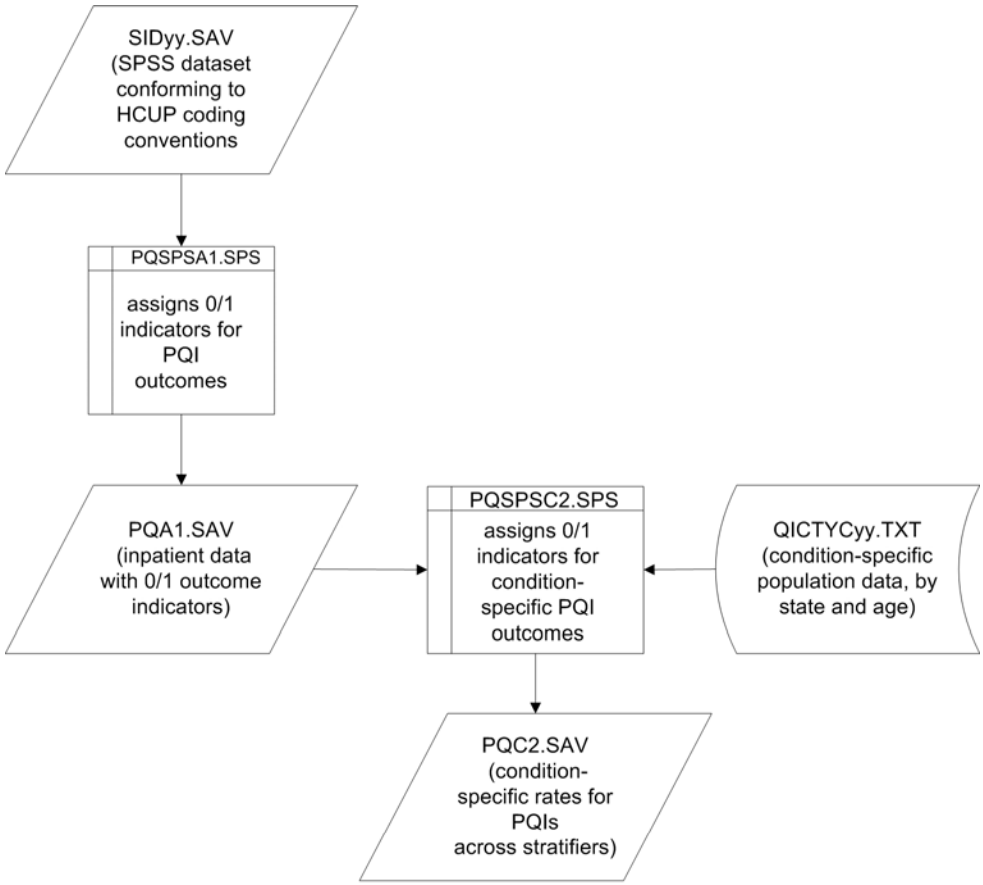
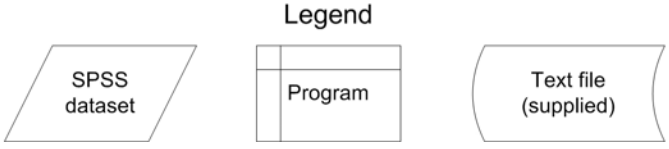


**Figure 2 Processing Steps for the PQIs - Condition-specific**



### *Prevention Quality Indicator Module Details*

## Table 1. Prevention Quality Indicator (PQI) Variables

| Indicator number | Numerator (admissions for ACSC) | Denominator | Age categories | | | |
|---|---|---|---|---|---|---|
| | | | 0 to 17 | 18 to 39 | 40 to 64 | 65 and over [+] |
| 1 | Diabetes short term cx | Area population | | Included | Included | Included* |
| 2 | Perforated appendix | **Appendicitis*** | Included | Included | Included | Included |
| 3 | Diabetes long term cx | Area population | | Included | Included | Included |
| 4 | Pediatric asthma | Area population | Included | | | |
| 5 | COPD | Area population | | Included | Included | Included |
| 6 | Pediatric gastroenteritis | Area population | Included | | | |
| 7 | Hypertension | Area population | | Included | Included | Included |
| 8 | CHF | Area population | | Included | Included | Included |
| 9 | Low birth weight | **Births*** | NA | NA | NA | NA |
| 10 | Dehydration | Area population | Included | Included | Included | Included |
| 11 | Bacterial pneumonia | Area population | Included | Included | Included | Included |
| 12 | Urinary infection | Area population | Included | Included | Included | Included |
| 13 | Angina w/o procedure | Area population | | Included | Included | Included |
| 14 | Diabetes uncontrolled | Area population | | Included | Included | Included |
| 15 | Adult asthma | Area population | | Included | Included | Included |
| 16 | Lower extremity amputation, DM | Area population | | Included | Included | Included |

ACSC – Ambulatory Care Sensitive Condition.
Cx – complications
CHF – Congestive Heart Failure
COPD – Chronic Obstructive Pulmonary Disease
DM – Diabetes Mellitus
* The denominator includes all hospital admissions for this condition in the area.
[+] The validity of these indicators for patients in this age category is less established, and should be considered separately from other age categories. This can be done by using age stratified rates.

## Table 2. Prefixes for the Prevention Quality Indicator (PQI) Variables

| Prefix | Contents | Example (for PQI # 1, Diabetes short–term complications) |
|---|---|---|
| T | Inpatient numerator (top) | TAPQ01 |
| P | Population denominator (pop) | PAPQ01 |
| O | Observed Rate | OAPQ01 |
| R | Risk-adjusted Rate | RAPQ01 |
| S | Smoothed Rate | SAPQ01 |
| C | Condition-specific Rate | CAPQ01 |

## Table 3. Prevention Quality Indicator Module Contents

| SPSS programs (syntax files) | ASCII text files (data) |
|---|---|
| PQSPSA1.SPS | QICTY**A**98.TXT |
| PQSPSA2.SPS | QICTY**A**99.TXT |
| PQSPSA3.SPS | QICTY**A**00.TXT |
| PQSPSC2.SPS‡ | QICTY**A**01.TXT |
| | QICTY**A**02.TXT |
| | QICTYA03.TXT |
| | QICTY**C97**.TXT |
| | QICTY**C98**.TXT |
| **ASCII text files (data)** † | QICTY**C99**.TXT |
| QICTY97.TXT | QICTY**C00**.TXT |
| QICTY98.TXT | QICTY**C01**.TXT |
| QICTY99.TXT | QICTY**C02**.TXT |
| QICTY00.TXT | QICTY**C03**.TXT |
| QICTY01.TXT | MNSPQA00.TXT |
| QICTY02.TXT | COVPQA00.TXT |
| QICTY03.TXT | MSXPQA_0.TXT |
| QICTY**A**97.TXT | |

† The ASCII text files are provided with the Prevention Quality Indicator module, and are necessary for the programs to run. There is no need to convert these files to SPSS format for use with the software. The parameter reference files of MSXPQA00.TXT and MSXPQA_0.TXT are not interchangeable between SAS and SPSS software. Users should verify they are running the software with the ASCII text files as listed in Table 3 of the applicable, e.g., SAS or SPSS, software manuals.

‡ Use to add condition-specific (diabetes) population figures and calculate rates by state and age.

## Background

The PQI module contains 16 indicators that measure hospital admissions for ambulatory care sensitive conditions (ACSC) across geographic areas. ACSCs represent conditions for which hospitalization could be avoided if the patient receives timely and adequate outpatient care. Many factors influence the quality of outpatient care, including access to care and adequately prescribed treatments, once care is obtained. In addition, patient compliance with those treatments and other patient factors may play a role. In general, areas with lower social-economic status tend to have higher admission rates for ACSCs than areas with higher social-economic status. As with utilization indicators, there are no "right rates" of admission for these conditions. Very low rates could signal inappropriate underutilization of healthcare resources while very high rates could indicate potential overuse of inpatient care.

Therefore, hospital admission for ACSCs is not a measure of hospital quality but a potential indicator of outpatient and community health care need. For example, if an area has a relatively high hospital admission rate for diabetes complications, the local healthcare providers should work with the community to identify reasons and strategies to address the problem.

The 16 ACSCs in the PQI module are listed in Table 1 on page 5. See Appendix A of the "*Guide to the Prevention Quality Indicators"* (http://www.qualityindicators.ahrq.gov/pqi_download.htm) for the definition of each indicator.

Except for perforated appendix and low birth weight, each indicator is measured as the number of hospital admissions for a particular ACSC per 100,000 residential population in an area. By default, the programs define areas by county; however, users have the option of specifying areas by Metropolitan Statistical Areas (MSAs) for urban counties. Rural areas are defined by county in either case.

The data required for measuring these indicators come from hospital discharge abstracts or billing claims (administrative data) which are readily available in many states. The residential population data are from the U.S. Census Bureau.

The software generates observed, risk-adjusted, and smoothed rates for each indicator at the area level. Observed rates are the raw rates. Risk-adjusted rates are derived from applying the average casemix of a baseline file that reflects a large proportion of the U.S. population to the areas' observed rates. Smoothed rates are estimates with removal of fluctuations over time due to random variation.

For information about how these indicators were identified, see the "*Guide to the Prevention Quality Indicators"* (http://www.qualityindicators.ahrq.gov/pqi_download.htm).

# Data Elements and Coding Conventions

## *Input for the AHRQ Quality Indicators*

The PQI software was written to process data from discharge data abstracts that contain information about hospital stays. The specific data elements needed for the PQI software are listed in Table 4. The PQI module was tested on data from the Healthcare Cost and Utilization Project (HCUP). HCUP is an ongoing Federal-State-private collaboration to build uniform databases from administrative hospital-based data collected by State data organizations and hospital associations.

*The input data files for the Prevention Quality Indicators software must be in SPSS.*

*We recommend that users recode data elements in their input files to be consistent with the coding expected by the software. This will minimize the number of internal changes that will be necessary in the Prevention Quality Indicators software. For example, recoding the SEX data element in the input file to be consistent with the coding described in Table 4 (i.e., 1 for male and 2 for female) is easier than modifying all uses of the SEX data element in the PQI programs.*

*Not every value for every data element is referenced by the PQI software. For example, admission source (ASOURCE) values are only used to identify transfers. However, we include the complete set of HCUP values to assist users of the uniform HCUP data files*.

*To minimize internal changes to the software, all required elements should be present in the input data file. If a required element is not available, a dummy element should be provided. Failure to provide a dummy element will result in errors during the execution of the PQI programs.*

The data elements listed in Table 4 are those required for the input data files of **all** the SAS and SPSS QI modules. When a variable is not required for the PQI SPSS program the notation "Not used by the PQI program" will be present in the Comments column. The SPSS program will still run if these variables are not present, however, the SAS program may not run correctly. Standardizing the variables and data values in the input data file will be helpful for users who have access to both statistical software packages or want to set up their input data file for use with any of the three QI modules: Inpatient Quality Indicators (IQIs), Prevention Quality Indicators (PQIs) or the Patient Safety Indicators (PSIs).

*Table 4 has been standardized to represent the input data file specifications required for use with any of the three QI modules (IQI, PQI or PSI). If a specific variable is not used by the PQI program it is not necessary to create a dummy variable in the input data file. For example, the variable APR-DRG is used only by the IQI software so it is not necessary to have this data element in the input data file or create a dummy variable for the PQI software program to run correctly.*

The data element HOSPSTCO provides flexibility for users to calculate the PQIs by hospital location or by patient residence. If the user wants to calculate the PQIs using the population associated with the hospital location as the denominator, the values for this variable should be the individual hospital FIPS[4] state/county codes. If the user wants to calculate PQIs based on the population of the MSA region or county associated with the patient residence, the values for this variable should be the FIPS state/county code associated with each individual patient's place of residence.

[4] Federal Information Processing Standard, as defined by the U.S. Department of Commerce, National Institute of Standards and Technology (formerly National Bureau of Standards).

*If the hospital FIPS code is used in HOSPSTCO, users should be aware that rates may be biased for hospitals which serve as regional referral centers. These hospitals are likely to treat patients from outside the MSA, county or even the state in which the facility is located.*

*Certain independent cities (Baltimore City, Carson City and St. Louis City), and areas within Hawaii and Virginia, are assigned to different area groupings in the Modified FIPS categories as compared to the US Census Bureau groupings. The AHRQ QI software uses the Modified FIPS code assignment of these areas. Failure to use the Modified FIPS codes assignment will result in the use of incorrect denominator estimates.*

*Using the patient FIPS state/county code for analysis may more accurately reflect the true population at risk. Evaluation of geographic variations in admissions for ambulatory care sensitive conditions by patient MSA or county of residence can result in better information to guide community or provider response.*

*It is possible that some records in the input data file may be missing the patient FIPS code. Users should be aware that any records with missing values (in the HOSPSTCO data field) are excluded from the calculations of observed, risk-adjusted and smoothed PQI rates. They will be included in the output from the first program (PQSPS1.SPS).*

*The software will generate rates for each county included in the HOSPSTCO data field. Users are encouraged to limit cases in their input file to those patients discharged from the geographic area of interest. For example, if you are using data from the state of Wisconsin and are interested in generating a rate for Wisconsin, you should remove the cases where the patient's county of residence (FIPS code) is from another state. Otherwise the total or overall rate will include the population of the counties outside Wisconsin.*

This version of the software provides the user with the option of producing output by MSA or by county. When MSA is selected, urban areas are defined by MSAs. When county is selected urban areas will be defined by county. Rural areas are always defined by county.

*In the software programs, !MSALEVL indicates whether MSA codes should be created from the FIPS state/county codes in the input dataset. If the user wishes to analyze data at the MSA level, this parameter should be set to 'yes'. If the user wishes to analyze data at the county level, this parameter should be set to 'no'.*

## *Coding of Diagnoses, Procedures, and DRGs*

*Diagnoses and procedures must be coded using the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM).*

*Significant modifications to ICD-9-CM codes occurred in the early 1990s. PQI definitions only reflect codes valid after October 1, 1994, and therefore may not accurately analyze data collected before 1994.*

- Diagnoses and procedure codes should be in character strings.
- Diagnosis and procedure codes should not contain any decimal points.
- Diagnosis and procedure codes should be left justified. Diagnosis codes with fewer than five digits should be padded with spaces (not zeros) on the right. Procedure codes with fewer than four digits should be padded with spaces (not zeros) on the right.

*The PQI programs assume accurate and specific coding of ICD codes. If an ICD code is specified using 5 digits, the software recognizes only 5-digit versions of that code and not versions truncated at 3 or 4 digits. Similarly, if an ICD code is specified using 4 digits, the software recognizes on 4-digit versions of that code and not versions truncated at 3 digits. For example, 4281 is the ICD-9-CM diagnosis code for left heart failure. If your data are coded less specifically (i.e., 428 is used to designate "heart failure"), these less specific codes are not recognized by the software and will result in undercounting cases.*

- This software has ICD-9-CM codes updated through FY 2005 changes.

- Diagnosis-related groups (DRGs) are those derived from the Centers for Medicare & Medicaid Services (previously Health Care Financing Administration) Medicare grouper. The software expects that you will be using the DRG effective on the discharge date. The software now refers to data elements DRG and MDC. Your data should be coded accordingly. The software may be modified at the user's option to use other types of DRGs. However, the impact of using other types of DRGs should be evaluated carefully before making such a change.

*Version 2.1, Revision 4, of the software accounts for ICD-9-CM and DRG coding changes effective through FY 2005 (or through September 30, 2005).*

## Table 4. Data Elements and Coding Conventions

| Variable name | Description | Format | Value description | Comments |
|---|---|---|---|---|
| KEY | Sequence number. Unique case identifier. | Numeric | User defined unique numeric identifier for each discharge record | Not used by the QI programs, but should be present to facilitate possible exploration; allows user to link the records from the PQSAS1.SAS program output file back to the original input data file. |
| AGE | Age in years at admission | Numeric | Age in years | If this data element is missing the discharge record will be excluded from analysis. |
| RACE | Race of patient. | Numeric<br>1<br>2<br>3<br>4<br>5<br>6 | White<br>Black<br>Hispanic<br>Asian or Pacific Island<br>Native American<br>Other | The values of 1, 2, and 3 are used directly in the QI software. All other ethnicity codes are mapped to an 'other' category. |
| SEX | Sex of patient | Numeric<br>1<br>2 | Male<br>Female | If this data element is missing the discharge record will be excluded from analysis. |
| PAY1 | Expected primary payer | Numeric<br>1<br>2<br>3<br>4<br>5<br>6 | Medicare<br>Medicaid<br>Private, incl. HMO<br>Self-pay<br>No charge<br>Other | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |
| HOSPSTCO | Location of patient residence or hospital location (FIPS† State/county code) | Numeric<br>ssccc | Modified Federal Information Processing Standards State/county code | Available at: http://www.census.gov/popest/geographic/codes02.pdf<br>If this data element is missing the discharge record will be excluded from rate calculations. |
| HOSPID | Data Source hospital number | Numeric | Hospital identification number | Used to facilitate data exploration and possible trouble shooting. |
| DISP | Disposition of patient | Numeric<br>1<br>2<br>3<br>4<br>5<br>6<br>7<br>20 | Routine<br>Short-term hospital<br>Skilled nursing facility<br>Intermediate care<br>Another type of facility<br>Home health care<br>Against medical advice<br>Died | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |
| ATYPE | Admission Type | Numeric<br>1<br>2<br>3<br>4<br>5<br>6 | Emergency<br>Urgent<br>Elective<br>Newborn<br>Trauma Center[5]<br>Other | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |

5 The ATYPE value "5" was used to indicate delivery only in the 1988 to 1997 HCUP data files. The UB standards currently use "5" to indicate Trauma Center.

| Variable name | Description | Format | Value description | Comments |
|---|---|---|---|---|
| ASOURCE | Admission Source | Numeric<br>1<br>2<br>3<br>4<br>5 | ER<br>Another hospital<br>Another fac. incl. LTC<br>court/law enforcement<br>routine/birth/other | The values 2 and 3 are referenced by the PQI code (to identify transfers from another hospital) |
| LOS | Length of Stay | Numeric | Number of days from admission to discharge | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |
| APR_DRG | APR-DRG category | Numeric | APR-DRG from 3M software. | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |
| SEVERTY | APR-DRG Severity Score | Numeric | APR-DRG Severity Score from 3M software. | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |
| RISKMORT | APR-DRG Mortality Score | Numeric | APR-DRG Risk of Mortality Score from 3M Software. | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |
| DRG | Diagnosis Related Group | Numeric | DRG from federal (CMS) DRG Grouper | If this data element is missing the discharge record will be excluded from analysis. |
| MDC | Major Diagnostic Category | Numeric | MDC from federal (CMS) DRG grouper | If this data element is missing the discharge record will be excluded from analysis. |
| DX1 – DX30 | Diagnoses ICD-9-CM codes. DX1 is the principal diagnosis, DX2-DX30 are secondary diagnoses. | String, 5 character | Diagnosis codes | Users with more or fewer secondary diagnoses should modify the parameter &NDX in CONTROL_PQI.SAS to reflect the number of diagnoses. |
| NDX | Count of the number of diagnoses (Dx) on the discharge record. | Numeric | Count of diagnoses (principal and all secondary diagnoses) | If the 1st discharge record has 5 Dx and the 2nd has 10, NDX would be 5 on the 1st and 10 on the 2nd. Not used by the SAS program. |
| PR1 – PR30 | Procedure ICD-9-CM codes. PR1 is the principal diagnosis, PR2-PR30 are secondary procedures. | String, 4 character | Procedure code | Users with more or fewer secondary procedures should modify the parameter &NPR in CONTROL_PQI.SAS to reflect the number of diagnoses. |
| NPR | Count of the number of procedures (Pr) on the discharge record. | Numeric | Count of procedures (principal and all secondary procedures) | If the 1st discharge record has 1 Pr and the 2nd has 2, NPR would be 1 for the 1st record and 2 for the 2nd. Not used by the SAS program. |
| PRDAY1 – PRDAY30 | Days from admission to procedure. PR1 is the principal procedure, PR2-PR30 are secondary procedures. | Numeric | Days from Admission to Procedure | Not used by the PQI program. If not present in the input data file it is **not** necessary to create a dummy variable. |

† Federal Information Processing Standard, as defined by the U.S. Department of Commerce, National Institute of Standards and Technology (formerly National Bureau of Standards).

Note: Certain independent cities (Baltimore City, Carson City and St. Louis City), and areas within Hawaii and Virginia, are assigned to different area groupings in the Modified FIPS categories as compared to the US Census Dept. groupings. The AHRQ QI software uses the Modified FIPS code assignment of these areas. Failure to use the Modified FIPS codes assignment will result in the use of incorrect denominator estimates for area indicators.

## *Missing Values*

*The PQI programs do not distinguish among different types of missing values.*

Data files of hospital discharge abstract data may have numeric data elements coded using special SPSS missing values. For example, besides the standard SPSS value of "." for missing data, there might also be present values of –1 for invalid data, -2 for data unavailable from a particular source, -3 for inconsistent data. However, the PQI programs do **not** distinguish among the different types of missing codes. Therefore, all types of missing values in the input data to the PQI module can be represented by a single value - missing data (.) for numeric variables and blank (' ') for alphanumeric (or character) variables.

*Cases that are missing data in fields used for indicator definitions or for risk-adjustment (such as age and sex) will be excluded from analysis.*

## *Treatment of Missing Data*

The software is designed to handle missing data in a specific manner; specifically, the software requires confirmation for the assignment of a poor outcome or negative event. For instance, in order to be excluded as a transfer, each case must actually be coded as a transfer. Missing data is considered neutral. Missing data for some elements results in the exclusion of that case from the numerator or denominator. For a few other elements, the case is retained. Table 5 lists the impact of missing data for each data element.

### Table 5. Treatment of Missing Data Elements

| Variable | Label | D | N | E | S | Treatment of Missing Data |
|---|---|---|---|---|---|---|
| AGE | Age In Years At Admission | X | X | | X | Case excluded from all analysis (e.g., dropped from numerator of all population based measures and from the denominator of the two discharge based indicators, PQI 2 and PQI 9) |
| ASOURCE | Admission Source | | | X | | Not counted in exclusion |
| DRG | Diagnosis Related Group (CMS DRG) in Effect on Discharge Date | X | X | | | Case excluded from all analysis (e.g., dropped from numerator of all population based measures and from the denominator of the two discharge based indicators, PQI 2 and PQI 9) |
| SEX | Patient Gender | X | X | | X | Case excluded from all analysis (e.g., dropped from numerator of all population based measures and from the denominator of the two discharge based indicators, PQI 2 and PQI 9) |
| HOSPSTCO | Location of Patient Residence or Hospital Location Modified FIPS State/County Code | X | | | X | Dropped from denominator in rate calculations (stratified by county or MSA), case will appear in calculation of the overall rate. |
| MDC | Major Diagnostic Category (CMS MDC) In Effect On Discharge Date | X | X | | | Case excluded from all analysis (e.g., dropped from numerator of all population based measures and from the denominator of the two discharge based indicators, PQI 2 and PQI 9) |
| RACE | Race | | | | X | Classified As "Other" |

D – Denominator; N – Numerator; E – Exclusion; S – Stratification

## User Options

The PQI software reflects the development and implementation of the software using data available to AHRQ. Our goal was to develop the tools, illustrate their use, and then encourage others to adopt and use the tools for their own applications. As a result, we expect and encourage users of the software to consider whether and how to modify the PQI software to better serve their local conditions or interests.

*Modifications to the definitions of outcomes of interest or populations at risk are possible but not desirable. Maintaining consistent definitions is important. Once definitions are altered, the ability to make comparisons of PQIs based on original definitions is lost. We encourage users to identify ways to improve the PQI methodology and to share their suggestions with us for future updates. Please contact us at support@qualityindicators.ahrq.gov.*

AHRQ has provided national estimates using the PQIs through its web-based query system, HCUPnet (http://hcup.ahrq.gov/HCUPnet.asp). If users change definitions of the PQIs, it will not be possible to compare users' results to the national estimates in HCUPnet.

## Processing Steps

Each Prevention Quality Indicator (PQI) expressed as a rate, is simply defined as:

Outcome of interest / Population at risk.

Conceptually, six steps are necessary to produce the PQI rates. The following describes the steps and how the software performs them.

### *1. Identify Outcomes in Inpatient Records*

Inpatient records are marked to indicate whether they contain the outcome of interest (numerator or "top") for each of the AHRQ PQI measures.

This is done by setting a series of flag variables, each of which corresponds to the numerator for a particular PQI. For example, if the inpatient record meets the conditions for inclusion in the outcome for PQI #1, then the outcome indicator for PQI #1 is set to 1.

This step requires one pass through the discharge-level data and outputs a discharge-level data file containing the original input variables and the flag indicator variables for the outcomes for each PQI.

### *2. Identify Populations at Risk from Census Population Data*

The populations at risk (the denominators for calculating the PQI rates) are derived from Census population figures by areas defined by county in all areas of the US, or by MSA in urban areas and by county in rural areas.

### *3. Calculate Observed (Raw) Prevention Quality Indicator Rates*

Using the output data from step 1 and Census population data from step 2, the PQI rates are calculated and saved for stratifiers which include user-defined combinations of area, age, race and sex.

*The programs calculate PQI rates regardless of the number of cases available. However, rates based on only a few cases should be interpreted with caution.*

In work at AHRQ, we do not report rates based on fewer than 30 cases in the numerator or the denominator. This exclusion rule serves two purposes:

- It eliminates unstable estimates based on too few cases.
- It helps protect the identities of hospitals and patients.

*Discharges with missing FIPS county codes are excluded in the output from step 3.*

### *4. Risk Adjust the Prevention Quality Indicator Rates*

Overall file means and regression coefficients from a baseline database (reflecting a large proportion of the U.S. population) are applied to the observed rates to risk-adjust the rates observed in the user's data. The risk-adjusted rates will then reflect the age and sex distribution of areas in the baseline file rather than the distribution for the areas covered by the users' data.

This will allow risk-adjusted rates produced by various users to be compared directly to one another. The overall means and regression coefficients were derived from AHRQ's State Inpatient Databases (SID) for 36 states. The codes to generate these baseline means and coefficients are not part of the PQI module.

### *5. Create MSX Smoothed Rates*

Shrinkage factors are applied to the risk-adjusted rates for each PQI in a process called multivariate signal extraction (MSX). These shrinkage factors were calculated from the SID database of 36 states. For each PQI, the shrinkage estimate reflects a 'reliability adjustment' unique to each indicator. The less reliable the PQI over time and across areas, the more the estimate 'shrinks' the PQI toward the overall area mean. The resulting rate will appear "smoother" than the raw rate, meaning the year-to-year fluctuations in performance are likely to be reduced. More information on interpreting smoothed rates is contained in the interpretation section of this document. The shrinkage factors are provided as part of the PQI software and do not need to be calculated by users.

### *6. Calculate Condition-specific Rates*

For selected indicators, data are available on the prevalence of the relevant condition. Version 2.1, Revision 4 incorporates state level estimates of diabetes prevalence by age from the CDC National Diabetes Surveillance System (www.cdc.gov/diabetes/statistics/index.htm). The four PQI related to diabetes (PQI #1, 3, 14 and 16) can be calculated using the number of diabetics in the state as the denominator, stratified by age.

These six steps reflect the PQI module production in a nutshell. The next section of this document describes the specifics of each software component of the PQI module software.

# Program Descriptions

This section describes the three SPSS programs that assign, calculate, and print the Prevention Quality Indicators.

For each program there is a description, a list of input and output files, and an explanation of changes to the program that may be required. The flow of data through the PQI module programs is shown in the flowcharts in Figure 1 and Figure 2 that begin on page 3.

If you desire to create and examine observed PQI rates, then the PQSPSA1.SPS and PQSPSA2.SPS programs will need to be used. If you also wish to create risk-adjusted and smoothed PQI rates, then you will also need to run the PQSPSA3.SPS program. If you wish to create condition-specific (diabetes) rates, you will need to run the PQSASC2 program.

## *Program 1: PQSPSA1.SPS*

The PQSPSA1.SPS program processes hospital discharge abstract data and flags inpatient records if they contain outcomes of interest. Outcome indicator names have prefix of "T" (Top). Stratifier variables are constructed at the beginning of the program.

This program assumes that the input data file (consisting of inpatient discharge abstract data) conforms to specific variable names, attributes, and coding conventions. See Table 4 on page 11 for variable names and attributes for the input data file.

Partial definitions of the Prevention Quality Indicators are given in Table 1 on page 5. This table is presented to assist those individuals who desire to examine the software source code statements. Complete definitions of the indicators are given in Appendix A of the "*Guide to the Prevention Quality Indicators".*

Input:

1. User supplied SPSS inpatient data set consisting of administrative hospital discharge abstract data. This data set is a discharge level file with an array of diagnosis and procedure codes, among other data elements.

Output:

1. SPSS data set (called PQA1.SAV) containing inpatient records with input variables, stratifiers, and flag indicators (TAPQxx) for the outcomes of interest that will later form the numerators for the PQI rates.

2. DESCRIPTIVES (with N, MIN, MAX, MEAN, and SUM) of all of the numeric variables in the output data file.

Changes:

Change the !LET parameters at the top of the program to specify the following:

1. !TEMPDIR should specify an existing directory where SPSS can save intermediate data sets temporarily. These data sets can be deleted once the program completes.

2. !PERMDIR should specify an existing directory where SPSS can save data sets with the final flags (PQA1.SAV). This is the directory where the program will look for the input data set with discharge records.

3. !TEXTDIR should specify the directory location where the AHRQ-supplied text files are stored.

4. !MSALEVL should indicate whether MSA codes should be created from the county codes (HOSPSTCO) on the input dataset. If the user wishes to analyze data at the MSA level, this parameter should be set to 'yes'. Specifying !MSALEVL='yes' does not affect the input HOSPSTCO codes, so users still have the option of analyzing data at the county level in the 2nd and 3rd program no matter how this parameter is specified in this first program.

5. !INDATA should specify the name of the input SPSS data set containing discharge records with procedures and diagnoses. This dataset must be located in the directory specified by !PERMDIR, above.

6. !MAXDX should specify the number of diagnosis variables in the input SPSS data set. By default, this value is set to 30.

 Users with fewer or more diagnosis codes should modify this value accordingly.

7. !MAXPR should specify the number of procedure variables in the input SPSS data set. By default, this value is set to 30.

 Users with fewer or more diagnosis codes should modify this value accordingly.

### *Program 2: PQSPSA2.SPS*

The PQSPSA2.SPS program calculates the observed or raw rates for the area-level Prevention Quality Indicators, using the data (PQA1.SAV) derived in the previous step (PQSPSA1.SPS). These observed rates are stratified by user-defined combinations of area, sex, age, and race categories. The program first totals the indicator flags created by the previous program, and then for each of the desired stratifiers divides these totals by the pertinent residential population. The population denominators are stored in variables with names that have a prefix of "P" (Pop). The Observed rates are stored in variables that have a prefix of "O".

Input:

1. The SPSS dataset created in Program 1 (PQA1.SAV).

2. A text file with Census area residential populations, stratified by area, age, sex, and ethnicity categories. Seven such files are currently provided along with the PQI module software. The files are QICTY97.TXT, QICTY98.TXT, QICTY99.TXT, QICTY00.TXT, QICTY01.TXT, QICTY02.TXT and QICTY03.TXT.

 The user should select the file for the year that best matches the user's input data set.

Output:

1. SPSS data set (PQA2.SAV) with summary records that contain observed rates (OAPQxx variables where xx refers to the indicator number), the counts of outcomes that formed the numerators of the rates (TAPQxx variables), and the residential population totals that formed the denominators of the observed rates (PAPQxx variables).

2. DESCRIPTIVES (N, MIN, MAX, MEAN, and SUM) of all of the numeric variables in the output data file.

3. An optional LISTING of the output summary dataset is provided at the end of the hardcopy printout. This printout may be quite large depending on the number and types of stratifiers that the user requested with the !STRLEVS parameter, discussed below. If the user does not want to generate this listing, the !PRINT parameter, discussed below, should be set to 'no'.

Changes:

Change the !LET parameters at the top of the program to specify the following:

1. !TEMPDIR should specify an existing directory where SPSS can save intermediate data sets temporarily. These data sets can be deleted once the program completes.

   *A pathname must end with the character "\" as shown in the example below:*

   !let !tempdir = 'c:\PSI\'

2. !PERMDIR should specify an existing directory where SPSS will read the data set from Program 1 (PQA1.SAV) and save the data set from this program (PQA2.SAV).

3. !TEXTDIR should specify the directory location where the AHRQ-supplied text files are stored.

4. !POPFILE should specify the name of the population file within the !TEXTDIR directory to be used in calculating the observed rates.

5. !MSALEVL should indicate whether observed rates should be constructed at the county level for all areas of the county (the default) or whether rates in urban areas should be constructed at the MSA level (!MSALEVL='yes').

   The specification of !MSALEVL must be consistent between programs 2 and 3.

6. !STRLEVS should specify the levels of stratification for which the program should calculate observed rates. The levels should be specified as a list of numbers (1 to 15 separated by commas) corresponding to the following stratifications shown in Table 6.

**Table 6. PQSPSA2 Stratification Choices**

| STRLEVS | Stratification | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | Race |
| 2 | | | | | Sex | | |
| 3 | | | | | Sex | * | Race |
| 4 | | | Age | | | | |
| 5 | | | Age | * | Sex | | |
| 6 | | | Age | * | | | Race |
| 7 | | | Age | * | Sex | * | Race |
| 8 | Area | | | | | | |
| 9 | Area | * | | | | | Race |
| 10 | Area | * | | | Sex | | |
| 11 | Area | * | | | Sex | * | Race |
| 12 | Area | * | Age | * | | | |
| 13 | Area | * | Age | * | | | Race |
| 14 | Area | * | Age | * | Sex | | |
| 15 | Area | * | Age | * | Sex | * | Race |

STRLEVS=0 (Overall) is calculated automatically.

Area will be defined differently depending on the specification of the !MSALEVL Parameter. If the parameter !MSALEVL is set to 'Yes', then urban areas will be defined by MSAs; if it is set to 'No', then urban areas will be defined by county. Rural areas will be defined by county in either case.

STRLEVS must include the value 8 to subsequently run program 3.

7. The !PRINT parameter should indicate whether the final dataset should be listed or not. If !PRINT is set to 'yes', then the program will generate a listing for each of the 16 indicators, including the numerator, denominator and observed rate for each level of stratification specified by the !STRLEVS parameter. If the user does not want to generate this listing, the !PRINT parameter should be set to 'no'.

## *Program 3: PQSPSA3.SPS*

The PQSPSA3.SPS program calculates age and sex risk-adjusted rates for each PQI (overall rates and rates by area), and then calculates smoothed rates.

Input:

1. The discharge-level dataset in SPSS that was created by Programs 1 (PQA1.SAV).

2. The SPSS dataset with summary records that was created by Program 2 (PQA2.SAV).

3. A text file with Census area residential populations, stratified by area, sex, and discrete age categories. Seven such files are currently provided along with the PQI module software. The files are QICTY**A**97.TXT, QICTY**A**98.TXT, QICTY**A**99.TXT, QICTY**A**00.TXT, QICTY**A**01.TXT, QICTYA02.TXT and QICTY**A**03.TXT.

The user should select the file for the year that best matches the user's input data file.

4. A text file (MNSPQA00.TXT) containing overall means from the reference SID dataset for use in the risk adjustment process. This file is provided to you as part of the PQI module. The text file should not be converted to SPSS.

5. A text file (COVPQA00.TXT) containing regression coefficients from a regression run on a reference SID dataset. These coefficients will be used in the risk adjustment process. This file is provided to you as part of the PQI module. The text file should not be converted to SPSS.

6. A text file (MSXPQA_0.TXT) containing three arrays for use in the smoothing process. The arrays contain noise estimates, signal estimates and mean area rates for each PQI. This file is provided to you as part of the PQI module. The text file should not be converted to SPSS.

Output:

1. SPSS data set (PQA3.SAV) containing the observed rates (OAPQxx variables), the risk-adjusted rates (RAPQxx variables), the smoothed rates (SAPQxx variables), the counts of outcomes that formed the numerators of the observed rates (TAPQxx variables), and the residential population totals that formed the denominators of the observed rates (PAPQxx variables).

2. DESCRIPTIVES (N, MIN, MAX, MEAN, and SUM) of all of the numeric variables in two intermediate work files and in the final output data file.

3. An optional LISTING of the output summary dataset is provided at the end of the hardcopy printout. If the user does not with to generate this printout, then the !PRINT parameter should be set to 'no'.

Changes:

Change the !LET parameters at the top of the program to specify the following:

1. !TEMPDIR should specify an existing directory where SPSS can save intermediate data sets temporarily. These data sets can be deleted once the program completes.

2. !PERMDIR should specify an existing directory where SPSS will read the data sets from Programs 1 (PQA1.SAV) and 2 (PQA2.SAV) and save the data set from this program (PQA3.SAV).

3. !TEXTDIR should specify the directory location where the AHRQ-supplied text files are stored.

4. !POPFILE should specify the name of the ASCII population text file to risk-adjust the rates. The file name will be QICTY**A**97.TXT, QICTY**A**98.TXT, QICTY**A**99.TXT, QICTY**A**00.TXT, QICTY**A**01.TXT, QICTYA02.TXT, or QICTY**A**03.TXT. These files are provided to you as part of the PQI module software. The user should select the file for the year that best matches the user's discharge data file.

5. !MSALEVL should indicate whether rates for urban counties should be calculated at the MSA rather than the county level.

   ⚠ !MSALEVL must be consistent between this program and the last.

6. !MEANSA should specify the name of the of the ASCII dataset that contains overall means from the reference SID dataset. The name of the file is MNSPQA00.TXT.

7. !COVARA should specify the name of the file containing the regression coefficients from a regression that was run on the reference SID dataset. The name of the file is COVPQA00.TXT.

8. !MSX should specify the ASCII file containing the estimates to smooth the risk-adjusted rates. The name of the file is MSXPQA_0.TXT.

9. The !PRINT parameter should indicate whether the final dataset should be listed or not. If !PRINT is set to 'yes', then the program will generate a listing for each of the 16 indicators, including the area designation (MSA or HOSPSTCO depending on how !MSALEVL was specified), numerator, denominator, observed rate, risk-adjusted rate and smoothed-area rate. If the user does not want to generate this listing, the !PRINT parameter should be set to 'no'.

## *Program 4: PQSPSC2.SPS*

The PQSPSC2 program calculates condition-specific rates for the four diabetes area-level Prevention Quality Indicators (PQIs 1, 3, 14, and 16), using the data (PQA1.SAV) derived in the previous step (PQSPSA1.SPS). These condition-specific rates are stratified by state and age categories. The program first totals the indicator flags created by the PQSPSA1 program, and then for each of the stratifiers divides these totals by the pertinent condition-specific population. The condition-specific denominators are stored in variables with names that have a prefix of "P" (Pop). The condition-specific rates are stored in variables that have a prefix of "C".

Input:

1. The SPSS dataset that was created in Program 1 (PQA1.SAV)

2. A text file with diabetes populations, stratified by state and age categories. Seven such files are currently provided along with the PQI module software. The files are QICTYC97, QICTYC98, QICTYC99, QICTYC00, QICTYC01, QICTYC02 and QICTYC03. The user should select the file for the year that best matches the user's discharge data file.

    *Users do **not** need to convert the ASCII text file to a SPSS dataset for use with the software.*

Output:

1. SPSS dataset with summary records that contain condition-specific rates (CAPQxx variables where xx refers to the indicator number), the counts of outcomes that formed the numerators of the rates (TAPQxx variables), and the condition-specific population totals that formed the denominators of the observed rates (PAPQxx variables). The output file has records for the overall state rate and by age.

2. DESCRIPTIVES (N, MIN, MAX, MEAN, and SUM) of all of the numeric variables in the output data file.

3. An optional LISTING of the output summary dataset is provided at the end of the hardcopy printout. If the user does not want to generate this listing, the !PRINT parameter, discussed below, should be set to 'no'.

Changes:

Change the !LET parameters at the top of the program to specify the following:

1. !TEMPDIR should specify an existing directory where SPSS can save intermediate data sets temporarily. These data sets can be deleted once the program completes.

2. !PERMDIR should specify an existing directory where SPSS will read the data set from Program 1 (PQA1.SAV) and save the data set from this program (PQC2.SAV).

3. !TEXTDIR should specify the directory location where the AHRQ-supplied text files are stored.

4. !POPFILE should specify the name of the population file within the !TEXTDIR directory to be used in calculating the condition-specific rates.

5. The !PRINT parameter should indicate whether the final dataset should be listed or not. If !PRINT is set to 'yes', then the program will generate a listing for each of the 4 indicators, including the numerator, denominator and condition-specific rate by state and state*age. If the user does not want to generate this listing, the !PRINT parameter should be set to 'no'.

## Reviewing the Printed Output

This section contains tips for reviewing some of the printed output from the PQI module. These tips are oriented toward explaining the interrelationships between printout items from different programs and hopefully will help to reveal the nature and structure of the module outputs. For guidance in interpreting the results, see the next section which begins on page 26.

### *PQSPSA1.SPS*

The initial printout from the PQSPSA1.SPS program contains descriptive statistics for all of the numeric variables in the output discharge-level dataset. It will contain information for the newly constructed TAPQxx flag variables that will later form the numerators for the indicator rates. For each TAPQxx flag variable:

- The SUM will contain the total number of observations in the dataset that were found to have the particular outcome of interest.

- For the majority (14 of 16) of the indicators, the MEAN, MINIMUM, and MAXIMUM will usually be the value one since the flag variables have either been set to missing ('.') or to a value of one. The two exceptions to this are the two indicators (#2, for Perforated appendix, and #9, for Low birth weight) that will based on subsets of the hospitalized population rather than the area residential population (see Table 1 on page 5). For these two indicators, a value of zero was assigned to the TAPQxx flag if a particular observation was part of the population for the rate denominator but did not have the particular outcome of interest to be included in the rate numerator. So for example, TAPQ02 = 0 implies a patient who had an appendectomy performed, but did not have a perforated appendix.

- For the two hospital-based indicators (#2 and #9), the MEANs will contain a close approximation of the eventual overall observed indicator rates. The values will change slightly after PQSPSA2.SPS has applied additional stratifiers such as area, age, sex and race/ethnicity.

- N lists the number of discharges in the dataset with non-missing values. For the 14 area-based indicators, N for TAPQxx will be the same as the SUM. For the two hospital-based indicators, N will contain the denominator for the observed indicator rate.

- There may be differences in the output from PQSPSA1.SPS and PQSPSA2.SPS and PQSPSA3.SPS programs based on missing data. If any cases are missing the FIPS codes they will be included in the output from PQSPSA1.SPS but will be excluded from the subsequent analyses (the second and third programs).

### *PQSPSA2.SPS*

The printout from the PQSPSA2.SPS program contains descriptive statistics for all of the numeric variables in the output summary dataset. It will contain information for the newly constructed OAPQxx rates, the PAPQxx denominators, and the TAPQxx numerators.

- The STRAT variable described in the first row of the table identifies the stratification level for the records in the output data set.

- The N statistic for STRAT contains the number of records in the output summary data set. A TAPQxx numerator variable with a lower value for N than STRAT indicates that there were no outcomes of interest in some of stratification cells. Similarly, a PAPQxx denominator variable with a lower value for N than STRAT indicates that for some stratification cells, the Census residential population estimate was zero.

- The MINIMUM value for the perforated appendix and the low birthweight TAPQxx numerators will be zero or one since values of zero were assigned for observations that were part of the population for the rate denominator but did not have the particular outcome of interest to be included in the rate numerator. For the other 14 indicators, based on the residential area population, the MINIMUM value will be one or higher.

- In general, the MEANS and SUMs in this printout have no intuitive meaning because numbers are added up repetitively over the stratifiers.

## *PQSPSA3.SPS*

*The **LISTING** at the end of this program provides your **final output**. (This printout appears if the "PRINT" parameter is set to 'yes'– the default setting.) It lists the numerator and the denominator, along with the observed, risk-adjusted, and smoothed rates for all indicators, as shown in the example output below.*

Diabetes Short Term Complication

List

| MSA | TAPQ01 | PAPQ01 | OAPQ01 | RAPQ01 | SAPQ01 |
|---|---|---|---|---|---|
| . | 7217 | 14989081 | .000481 | .000479 | .000479 |
| 40.00 | 88 | 90808 | .000969 | .000965 | .000898 |
| 320.00 | 77 | 157162 | .000490 | .000490 | .000486 |
| 640.00 | 348 | 916415 | .000380 | .000372 | .000373 |
| 840.00 | 166 | 282702 | .000587 | .000590 | .000583 |
| 1240.00 | 94 | 220132 | .000427 | .000426 | .000426 |
| 1260.00 | 60 | 112061 | .000535 | .000511 | .000503 |
| 1880.00 | 175 | 265449 | .000659 | .000659 | .000648 |
| 1922.00 | 1879 | 3810513 | .000493 | .000490 | .000490 |
| 2320.00 | 278 | 462274 | .000601 | .000599 | .000595 |
| 3362.00 | 1791 | 3342014 | .000536 | .000534 | .000533 |
| 3810.00 | 69 | 222232 | .000310 | .000302 | .000310 |
| 4080.00 | 62 | 123899 | .000500 | .000494 | .000489 |
| 4420.00 | 105 | 151799 | .000692 | .000693 | .000672 |
| 4600.00 | 109 | 175436 | .000621 | .000615 | .000603 |
| 4880.00 | 161 | 365898 | .000440 | .000437 | .000437 |
| 5800.00 | 97 | 163002 | .000595 | .000597 | .000585 |

Observation 1 (with MSA = .) is the overall average for the entire dataset (STRAT = 0). The remaining observations are individual MSAs (STRAT = 8).

You may wish to express the results in more understandable terms:

- Multiply the area rates by 100,000 to express them as a rate per 100,000 population (e.g., 0.000479 * 100,000 = 47.9 cases of Diabetes with Short Term Complications per 100,000 population).

The processing performed by this program is primarily at the area level. (Overall statistics for your dataset are also produced.) As a result, the PQSPSA3.SPS printed output is easier to interpret than the output from the preceding run when multiple stratifications may be in play.

One call to descriptives is run on the permanent area-level output file.

- The N should contain the number of different areas (MSAs and counties) in your database.

- The MEANs, MINIMUMs, and MAXIMUMs have their normal meaning and provide comparisons among the different areas in your database. Note that the MAXIMUMs for the counter variables (the TAPQxx and PAPQxx variables) are associated with specific areas, and therefore these MAXIMUMs

may not match those in the prior PQSPSA2.SPS printouts since that run typically will include a record for the entire database.

- The SUMs of the counter variables (the TAPQxx and PAPQxx variables) yield the overall database totals, which could be different from those in the output of PQSPSA1.SPS.

- Users should note the totals for the population based measures may be different based on the area stratifier. In the county-level stratification, a county population is included only if it has a discharge. Counties with no discharges are not included. In the MSA-level stratification, a county is included if any county within the same MSA had a discharge. So counties with no discharges can be included. Therefore the overall populations may be bigger and the overall rates smaller when stratifying by MSA.

## *PQSPSC2.SPS*

The printout from the PQSPSC2.SPS program contains descriptive statistics for all of the numeric variables in the output summary dataset. It will contain information for the newly constructed CAPQxx rates, the PAPQxx denominators, and the TAPQxx numerators.

- The STRAT variable described in the first row of the table identifies the stratification level for the records in the output data set.

- The N statistic contains the number of states in the output summary data set.

- The MINIMUM value will be one or higher (for PAPQxx and TAPQxx) or zero or higher (for CAPQxx).

- In general, the MEANS and SUMs in this printout have no intuitive meaning because numbers are added up repetitively over the stratifiers.

## Interpreting the Results

After performing all of the steps as previously outlined, users will have three estimates of performance for each PQI – observed, risk-adjusted, and MSX smoothed. This section will outline how these estimates can be used.

The Prevention Quality Indicators are not intended as definitive quality measures. But they are useful, low-cost measures that can potentially illuminate differences across areas in hospital admissions for ambulatory care sensitive conditions (ACSC).

Since there are no "right rates" established for most indicators, it is often best to compare area-level rates with other similar areas. These "peer groups" would ideally be as similar as possible in potentially important factors, such as socioeconomic status of the population, and urban or rural location.

Performance on a single PQI often cannot reliably indicate actual quality differences. For this reason, some indicators have been developed as measure sets. For instance, diabetes has four different ACSC indicators – uncontrolled diabetes, diabetes short-term complications, diabetes long-term complications, and lower-extremity amputation among patients with diabetes. Examining these indicators together is likely to produce a more complete picture of overall quality of care for this condition.

### *Observed Means*

Interpretation of the observed (raw) PQI rates may be complicated by concerns over noise and bias. However, the observed rates do not require complex statistical methods as do the risk-adjusted and MSX smoothed rates. For this reason, they are the most intuitive estimates of PQI performance. Observed rates can be used as a baseline measure when comparing to risk-adjusted and MSX smoothed rates, to determine the impact of risk adjustment/smoothing.

### *Comparing Observed and Condition-Specific Means*

The condition-specific rate will be equal to the observed rate multiplied by (1 / prevalence rate). For example, if the observed rate is 200 per 100,000, and the prevalence rate is 1,000 per 100,000, then the condition-specific rate is (200 / 100,000) * (100,000 / 1,000) or 200 per 1,000. In general, the condition-specific rate is more informative because it tells the preventable admission rate among the population at risk for that condition. However, the observed rate provides perspective on the frequency of admission across conditions with varying prevalence.

### *Risk-adjusted Rates*

Simply put, risk-adjusted rates are the estimated performance of areas on the PQIs if those areas had an 'average' casemix, as defined by age and sex. This average casemix is estimated and included in the software using estimates from 36 states in AHRQ's State Inpatient Databases (SID). Estimates of the average casemix do not reflect only the areas in the dataset being analyzed, but rather the distribution in age and sex among the areas in the SID data.

All area indicators are risk-adjusted using age and sex, except for low-birth weight, which is not risk-adjusted. Ideally, one might like to adjust for the underlying health status of each area, but population-based measures of health status by demographic groups are not readily available.

Users should compare risk-adjusted rates with the observed rates. This will indicate the impact of risk adjustment and presumably casemix on indicator performance. Areas with large changes in performance between observed and risk-adjusted rates presumably have a more or less complex case mix.

The construction of risk-adjusted rates in the PQI module follows the method described in Greene, Econometric Analysis (2nd Ed), 1993, Section 16.4.2 on calculating fixed effects in a linear model. Two

important features of this approach should be noted. First, the risk-adjusted rates are coefficients on area level covariates ('fixed effects') in a linear model for each PQI that also accounts for patient demographics (age and sex). Generally, these coefficients reflect the relative performance on the PQI relative to a reference or 'omitted' group (e.g., males age 65-69). As a result, the risk-adjusted rates for each area would reflect performance on the PQI for the reference group only, making interpretation and comparison with the observed raw rates problematic. Rather, the approach incorporated into running the model is to first subtract the overall PQI mean for each covariate (e.g., age category 1, age category 2, etc.) prior to running the model. The benefit of this approach is that the risk-adjusted rates (the 'fixed effects' in the model) now reflect the estimated performance on the PQI for each area if that area had the 'average' casemix among all areas in the estimation sample.

Second, because the model is linear, rather than logistic, the estimated risk-adjusted rates may be negative (that is, the model does not require the rates to be between zero and one). This might happen, for example, if the observed rate is close to zero, and the area has a more severe than average casemix. The model would predict that the area's performance with an 'average' casemix would be less and subtract some adjusted amount from the observed rate. If that observed rate is already nearly zero, the adjusted rate may be negative. A negative rate means that the observed rate is close to zero, and that the area has a less-severe-than-average casemix.

In order to facilitate the interpretation of the estimates, the PQI module imposes a lower bound of zero; i.e., negative values are set to zero.

### *MSX Smoothed Rates (Optional)*

The PQSPSA3.SPS program also yields the estimated MSX smoothed rates for each PQI. These smoothed rates are estimated using multivariate signal extraction (MSX). MSX is a Bayesian smoothing technique. Generally, indicators defined on relatively small populations or for relatively rare events are very noisy measures. In other words, many other factors other than quality can influence the observed rate. For factors that we can observe, such as the patient's age or gender, risk adjustment can account for their influence. There are many other clinical and non-clinical factors that we cannot observe. However, the MSX method estimates how much of an impact random differences in these factors have on the observed rate. The program accomplishes this by adjusting toward the overall area mean according to the degree of reliability of the specific area indicator.

Just as risk-adjustment levels the playing field across areas by "equalizing" population characteristics, smoothing levels the playing field by removing random variation in the measures across areas and over time. In essence, smoothing tells us how persistent we would expect an area's rate to be from year to year.

The smoothed rates are intended to help you look for how best to present the data. For example, if smoothed rates vary considerably from risk-adjusted rates, you might wish to examine risk-adjusted performance by combining data for several years rather than examining a single year's performance, or just compare similar areas.

The MSX method estimates reliability by looking at area performance among multiple indicators and across multiple years for a single indicator. If areas that perform high in one year also tend to perform high in the next, and areas that are low in one indicator tend to perform low in other related indicators, then the methods incorporate this information in forming an estimate of how much of an area's performance on a single indicator in a single year is "signal" and how much is "noise". PQSPSA3.SPS reports the "signal" estimate, after removing the "noise".

These estimates are referred to as "smoothed" estimates because when the raw data is reported annually on a graph, for example, the line connecting the points may show massive fluctuations, seemingly with no apparent pattern. Some of these fluctuations are due to random differences in factors that influence performance, so once the impact of these factors is removed, fluctuations are reduced, or smoothed out.

Improvements in performance will still be observed, however, they will only be observed when it is likely that these improvements are actually true differences, and not simply random variation.

The estimates of how related the Prevention Quality Indicators are among measures and over time for a typical area are based on estimates using the 36 states in the SID database. Therefore, PQSPSA3.SPS does not require multiple areas, measures, and years of data to operate. Instead shrinkage estimates were calculated from the combination of 27 statewide HCUP data sets to which the MSX had been applied. These shrinkage estimates are the best estimates of the impact that MSX smoothing would have on a particular PQI if MSX could be applied to the data set being analyzed. These shrinkage estimates are then applied to the adjusted rates for each PQI, giving 'smoothed' estimates of the rates.

The smoothed estimates represent the best guess of what the area performance would be for any given PQI if the data contained only a minimal amount of noise. This has several implications. First, users can be more certain that changes observed over time represent true differences in performance, rather than random variation. Second, in any given year, users can be more certain that the rate does not reflect only random noise. This can best be seen by comparing the smoothed estimates with the observed raw rates, which will indicate the impact of MSX on that indicator. A large impact means that the observed rate of the indicator is less precise, meaning it contains more noise relative to other indicators sustaining less impact. Low precision can result from small numbers of observations, or for relatively rare events. Therefore, the impact of the MSX estimate will in general be greater for infrequent events. Plotting the MSX estimates over time will reflect more persistent performance from year-to-year, and the correlations among indicators will generally be stronger.

## *Comparing Observed, Risk-adjusted and Smoothed Rates*

The purpose of the analysis determines which rates the user should look at in evaluating the performance of an area. If the user's primary interest is to focus on a particular area without any comparisons to other areas, simply examine the overall observed rate for the entire area, as well as further breakdowns by age, sex, and race/ethnicity which are the output created by the PQSPSA2.SPS program. Trend data are recommended for identifying changes over years.

If the purpose of the analysis is to compare the performance of a particular area with national, state, or regional averages or performances of other selected areas, then both observed and risk-adjusted rates should be examined. Variation in observed rates across areas is attributable to a variety of factors including differences in population demographics, disparity in access to and quality of care or other area characteristics ('systematic factors'), and random factors (non-systematic factors or 'noise'). Comparing observed and risk-adjusted rates can reveal if there is any difference between the area's population and the population of other areas. If the difference is minimal, one can compare the area's observed rate with the overall average across all areas. However, to account for differences in population demographics among different areas, risk-adjusted rates should be used for area-by-area comparisons.

After removing the impact of population demographics, if users want to examine potential existence of random factors, comparisons can be made between the risk-adjusted rates and the smoothed rates to determine if the differences in risk-adjusted rates across areas are due to systematic, as opposed to random, factors. The following two tables provide guidance on how to interpret differences in observed, risk-adjusted, and smoothed rates.

## Table 7. Comparing Observed Rates with Risk-adjusted Rates

*Purpose:* *To identify if there is any difference in the demographic composition of the area relative to the demographic composition of all areas combined*

| | |
|---|---|
| Observed rate **>** Risk-adjusted rate | The area's population has a ***higher*** risk of being admitted to hospitals for the condition due to its demographic composition (for example, older or a greater proportion of a higher-risk gender). |
| Observed rate **<** Risk-adjusted rate | The area's population has a ***lower*** risk of being admitted to hospitals for the condition due to its demographic composition (for example, younger or a greater proportion of a lower-risk gender). |
| Observed rate **=** Risk-adjusted rate | The area's population is similar to other areas in demographic composition, suggesting that the demographic composition is not a contributing factor to the area's performance for the indicator. |

## Table 8. Comparing Risk-adjusted Rates with Smoothed Rates (Optional)

*Purpose:* *To determine if differences in risk-adjusted rates across areas are due to systematic or random factors after demographic composition is taken into account.*

| | |
|---|---|
| Risk-adjusted rate ***similar to*** Smoothed rate | The area's risk-adjusted rate is more precise, an accurate reflection of the area's systematic ('true') rate. If there is any difference between the risk-adjusted rate and the overall area average[1], such difference is likely to be due to systematic factors. |
| Risk-adjusted rate ***different from*** Smoothed rate | The area's risk-adjusted rate is less precise[2]; that is, some of the difference between the risk-adjusted rate and the overall area average is due to random ('noise') factors. The smoothed rate is thus a more accurate estimate of the systematic rate than the risk-adjusted rate. Comparing the smoothed rate to the overall area average reveals more accurately the 'true' performance of the area. |

Notes: 1. The overall area average refers to the weighted average risk-adjusted rates across all areas.
2. The risk-adjusted rates for areas with small population are generally less precise.

Generally, subtracting the observed, risk-adjusted and smoothed rates can help to reveal differences among these rates. To get a sense of the magnitude, the most straightforward approach is to estimate the difference in terms of the number of cases in the numerator. For example, for an area with 35,000 pediatric population (under 18 years of age), a difference of 0.0008 between the observed rate and the risk-adjusted rate would mean 28 hospital admissions. Whether the difference between rates is statistically meaningful will depend on the size of the denominator and the sampling variability for each indicator. Users can apply a rough rule of thumb to estimate the required difference: 1.4* SQRT(p*(1-p)/N), where p is the rate and N is the denominator of the indicator.

For example, if the observed rate for an indicator is p=0.0045 and the denominator is N=35,000, applying the formula to these numbers will produce: 1.4* SQRT(0.0045*(1- 0.0045)/35,000) = 0.0005. That is, for the risk-adjusted rate to be significantly different from the observed rate, the absolute difference between the two rates has to be equal to or greater than 0.0005.

## Benchmark Timings

The benchmark runtimes given below are from runs made on a 500 MHz Pentium III, running Windows 98, with 256 MB of RAM, and an IDE hard drive. The dataset used contained 100,000 observations with 15 diagnosis fields and 15 procedure fields.

| Step | Run time (in seconds) |
|---|---|
| 1. PQSPSA1.SPS | 050 |
| 2. PQSPSA2.SPS | 029 |
| 3. PQSPSA3.SPS | 220 |

## User Support and Information

We would like to receive your feedback on the AHRQ Quality Indicators.

Our E-mail address for user feedback is: support@qualityindicators.ahrq.gov

We offer a listserv to keep you informed on the Quality Indicators. We encourage you to sign up for this free service. All you need is a computer, Internet access, and an E-mail address. It works just like other electronic distribution lists.

Here's how to register:

1. Send an E-mail message to: listserv@qualityindicators.ahrq.gov.
2. On the subject line, type: Subscribe. For example:

   Subscribe
3. In the body of the message type: sub Quality_Indicators-L and your full name. For example:

   sub Quality_Indicators-L John Doe
4. You will receive a message confirming that you are signed up.

If you have any questions, contact AHRQ QI Support at the e-mail noted above. If you do not receive a confirmation message call (888) 512-6090.

# Appendix A: Log of Revisions to PQI Documentation and Software Version 2.1, Revision 4

The following table summarizes the revisions made to the PQI software, software documentation and the Guide to Patient Safety Indicators (Guide) document in release version 2.1, revision 4. The table lists the component(s) affected by the change and a short summary of the changes that were made.

| ***Component*** | ***Changes*** |
|---|---|
| Software (SAS and SPSS), Software Documentation, and Guide | 1. The years for which the ICD-9-CM and DRG codes defining PQIs are valid was amended to be through FY 2005 instead of FY 2004, that is, the codes in the software are effective through September 30, 2005.<br>2. Added new module that calculates condition-specific rates for PQIs across stratifiers. |
| Software Documentation (SAS and SPSS) | Table 3 was amended to include the 2003 census data (i.e., QICTY03.TXT and QICTYA03.TXT) and condition-specific module files (PQSASC2 and QICTYC03.TXT). |
| Software (SAS and SPSS) | Added the 2003 census data (i.e., QICTY03.TXT and QICTYA03.TXT) and condition-specific module files (PQSASC2 and QICTYC03.TXT). |
| Guide | Rearranged the sequence of PQIs to place in numerical order. |
| Software (SAS) | Inserted “PQ” in format names for age aggregations in SAS programs to distinguish these formats from similarly named formats used by other indicator software. |

## Appendix B: ICD-9-CM and DRG Coding Updates in PQI Release Version 2.1, Revision 4

For Fiscal Year 2005 (effective 10-1-2004) there were no ICD-9-CM or DRG coding changes that affected indicator definitions. Refer to the Prevention Quality Indicators Archive (http://www.qualityindicators.ahrq.gov/pqi_archive.htm) for changes to indicator definitions made in previous PQI revisions.